Editorial

 doi:10.1093/bfgp/ell016

Editorial

THE RICH DIVERSITY OF GENOMICS—A REPORT ON THE 'COMPARATIVE AND FUNCTIONAL GENOMICS (BITS) WORKSHOP', HINXTON, UK, 27–30 OCTOBER 2005

The Comparative and Functional Genomics (BITS) Workshop has a history that in many ways reflects the changing face of modern genomics. Started 15 years ago under the banner of 'Identification of the Transcribed Sequence', the meeting was designed to bring together leading researchers from around the world who were pioneering new global approaches to gene discovery in a small workshop setting. As more and more transcribed sequences became known, the emphasis of the meeting, like the community it served, focused on how to characterize the function of all the newly acquired genes. A decision was therefore made to change its name to 'Beyond the Identification of the Transcribed Sequence Workshop', or BITS for short. As the years have passed the meeting has continued to diversify and change (as has its name), but it has continued to attract scientists to the cutting edge of genomics research. At this year's meeting, hosted for the second time at the Wellcome Trust Genome Campus, Hinxton, UK, around 80 participants were treated to a packed programme of over 30 presentations. These covered an eclectic range of current topics in genomics from new technologies, resource creation, computational biology, functional genomics, systems biology and various combinations of the above. This edition of the BFGP journal has aimed to capture the flavour of this diversity with pieces written by contributors to the meeting.

Reflecting its roots, the meeting has always retained progress reports from groups undertaking large-scale cDNA sequencing programmes. However, in recent times the focus of their talks has moved increasingly on to their efforts to use and characterize the resources they have created. Osamu Ohara (Kazusa DNA Research Institute, Japan) described their efforts to characterize genes in the Kazusa mammalian cDNA resource (KIAA gene clones) and Stefan Wiemann (DKFZ, Heidelberg, Germany), the ORFeome pipeline they have developed to screen for and characterize candidate genes that play a role in cancer progression, particularly those that have been involved in sequencing.

The International Mouse Knock-out Project is a co-ordinated worldwide effort to generate mutations in every gene in the mouse genome. Two of the leading groups in this consortium were represented at the workshop. Geoffrey Hicks (University of Manitoba, Winnipeg, Canada) and Frank Schnütgen (University of Frankfurt Medical School, Germany) presented updates on the North American and European Conditional Mouse Mutagenesis programmes (NorCOMM and EUCOMM) including new technologies, number of lines generated, their characterization and availability, together with their own studies on mice, generated as part of this programme. Use of mutagenesis as a tool to discover gene function was also described in other species. Adam Amsterdam (MIT Centre for Cancer Research, Cambridge, USA) described their use of genome-wide zebrafish mutant collections to discover and characterize genes involved in development and to study their potential role in cancer.

Studies into various aspects of the control of gene expression were also presented. Thomas Cremer (Ludwig–Maximilians–Universität München, Martinsried, Germany) described their work on establishing fundamental principles of higher order chromatin arrangements and the effect that topological modifications to chromatin have on regulating cell function. The role of chromatin in influencing transcription was also explored by Juergen Bode (German Research Centre for Epigenetic Regulation, Braunschweig) who described their studies on the interferon-beta promoter and the identification of novel long-distance regulatory domains. Thomas Werner's (Genomatix, Munich, Germany) presentation discussed how they were combining the results from high-throughput

genomics platforms with the company's *in silico* platform to accurately predict regulatory domains in the upstream regions of genes.

RNA editing is now widely recognized as another potentially important level of control. Katheleen Gardiner (Eleanor Roosevelt Institute, Denver, US) described their studies on the A-to-I editing of the 5HT2C receptor mRNA and their efforts to understand the functional consequences of these changes, in particular with respect to behavioural differences that may result from this type of RNA editing. A-to-I RNA editing was also discussed by Eli Eisenberg (Tel Aviv University, Tel Aviv, Israel), with their focus being to identify and understand the apparent differences in A-to-I editing between human and mouse. Work on alternative splicing, another common form of mRNA modification was also presented at the meeting. Graziano Pesole (University of Milan, Italy) described their computational methods and databases for the prediction of alternative splicing events, and Tom Blummenthal (University of Colorado, US) their studies on the unusual characteristics of this important phenomenon in *Caenorhabditis elegans*. In a similar vein, Stefan Stamm (University of Erlangen, Germany) discussed their studies on the signal transduction pathways that control alternative pre-mRNA splicing.

Two talks were presented from groups performing large-scale RNAi screens. Andrew Fraser (Sanger Institute, Hinxton, UK) gave an update on their work of RNAi screens on various mutants of *C. elegans*. Phenotypic changes in the worms can be used to suggest pathways in which the genes might be acting. Similarly, Michael Boutros (DKFZ, Heidelberg, Germany) reported on their gene knock-down studies in *Drosophila*. With genome-wide RNAi libraries now becoming available for mammalian genomes, reports on the application of this technology are likely to be a prominent feature of future meetings.

Comparative genomics have been used by Steven Brenner, who described an application of Bayesian phylogenomics through construction of a statistical graphical model of evolutionary relationships to predict functions of uncharacterized proteins. Greg Elgar (University of London, UK) discussed how they had used comparative genomics to identify highly conserved non-coding sequences present in all vertebrate genomes. These sequences are closely associated with key genes that orchestrate early vertebrate development and he described an assay in zebrafish embryos, which allows one to test these conserved non-coding elements (CNEs) for enhancer activity *in vivo*.

Using genomic tiling path arrays Tom Gingeras (Affymetrix, Santa Clara, US) and his group were instrumental in unveiling the previously little studied world of non-protein coding transcripts that are clearly now an important component of the transcriptome. He updated the meeting on their explorations into these transcripts of unknown function (TUFs) and their insights into the possible regulatory role of these abundant, but poorly understood transcripts. Perhaps we are not 'beyond' identifying all the transcribed sequences after all! The data produced by tiling arrays comes with its own challenges. Wolfgang Huber (EBI, Hinxton, UK) presented their work on analysing such data, and in particular on the adjustment of probe sequence effects by using calibration data from genomic DNA hybridizations as a control, and on a method for mapping of transcript boundaries, levels and internal architecture through a dynamic programming change point detection algorithm. He also described a series of novel findings on ncRNA, including isolated transcripts, antisense transcripts and untranslated regions (UTRs) of coding genes that were emerging from the use of these arrays in *Saccharomyces cerevisiae*.

Progress in the field of protein arrays was presented by Marcus Templin (University of Tübingen, Reutlingen, Germany). He described their efforts in establishing different protein array systems capable of detecting and quantifying marker proteins directly from serum, cell culture and tumour material. John McCafferty (Sanger Institute, Hinxton, UK) described his groups' progress on what must surely be one of the most ambitious proteomics projects around today. They have been optimizing the production of recombinant antibodies with the long-term aim of producing an antibody to every human and mouse protein. These are then being used for immunohistochemical studies to produce an 'Atlas of Gene Expression' describing the localization of proteins across a broad spectrum of tissues.

As an increasing number of high-throughput functional genomics datasets are generated, many are now attempting to integrate and analyse these data in various ways, for instance trying to infer biological networks. Winston Hide (SANBI, University of

Western Cape, South Africa) described their efforts in developing a formal ontology (eVOC) that is allowing them to combine data from information resources on gene expression and promoter analysis to elucidate key aspects of regulatory networks. Ravi Iyengar (Mount Sinai School of Medicine, New York, US) described their studies using graph theory to characterize the regulatory patterns that are formed as receptor signals propagated through networks representing the mammalian hippocampal neuron. Alvis Brazma (EBI, Hinxton, UK) described their work on modelling gene networks using public datasets for yeast mutants and array-based chromatin immunoprecipitation experiments. Efforts to build up networks of interacting proteins, the 'Interactome', were described by Michael Cusick (Harvard Medical School, Boston, US). Their group have for a number of years, been performing high-throughput yeast-2-hybrid screens and are now beginning to integrate these data with data from other sources in an effort to interpret protein networks. Finally, Peter Karp (SRI International, Menlo Park, US) described the BioCyc database collection. BioCyc is a set of 160 pathway and genome databases for most of the prokaryotic and eukaryotic organisms whose genomes have been completely sequenced.

Finally, to those speakers difficult to pigeonhole as belonging to one genomics field or another due to the use of a range of tools in their investigations. The group of Kevin White have, for a number of years, been taking the lead in exploring *Drosophila* biology. He described not only the impact that these studies were having on the understanding of fly biology, but also how these findings were being translated into insights into human disease. Likewise, Christian Maerker (RZPD, Heidelberg, Germany) has been using a variety of tools to explore the underlying causes of ventricular hypertrophy and Sherman Weissman (Yale University, US) to explore cell differentiation pathways in the haematopoietic system. Last but by no means least, Seth Grant (Sanger Institute, Hinxton, UK) presented their use of genomic technologies to unravel the exquisite complexity of protein networks of the mammalian synapse.

We hope that this brief introduction will provide a better idea of the meeting beyond what you might be able to ascertain from the following articles. On a final note, we would just like to make a case for this kind of a meeting. Its small participant numbers, egalitarian ethos, and broad and relaxed programme provide an ideal opportunity not only to hear about a wide range of great scientific research, but also to talk to the speakers and have the time to think about its relationships to ones own work. Unfortunately, as the local organizers of this year's meeting, we are only too well aware that underwriting such events where financial 'losses' are always guaranteed, is never easy. However, with numerous participants returning for several years in succession and enthusiasm for continuation of this meeting being high, all will be done to ensure that it remains an important part of the scientific calendar.

The Comparative and Functional Genomics (BITS) Workshop was funded by Wellcome Trust, UK and the Department of Energy, US.

Tom Freeman
Scottish Centre for Genomic Technology and Informatics, University of Edinburgh Medical School, The Chancellor's Building, Edinburgh EH16 4SB, UK

Alvis Brazma
European Bioinformatics Institute, EMBL-EBI, The Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK